

# **An Application of Biological Information in a Compendium**

---

Students: Li McCarthy, Pengcheng Wu  
Blackboard Group Name: McCarthyWu

March 1, 2019

## CONTENTS

1 Introduction	3
1.1 Top-Level Description	3
1.2 Functionality to the User	3
1.3 Storage and Logistics	3
2 Conceptual Database Design	4
3 Activity Diagram	5

# 1 INTRODUCTION

## 1.1 TOP-LEVEL DESCRIPTION

In biology field, a perturbation is any change in gene expression which can result from genetic disturbance, small molecule activity, or disease. The CMap, or “Connectivity Map”. (<https://www.broadinstitute.org/connectivity-map-cmap>), is a large-scale dataset maintained by the Broad institute which documents and cross-compares these perturbations. It facilitates the discovery of connections between genes, drugs, and diseases. The cloud-based data storage termed CLUE (<https://clue.io/>), is a integrated database environment also maintained by the institute, allowing users to query differences in gene transcription, which is based on 9 core cell lines across around 30 tissues. Computational biologists can access data using data APIs at <https://clue.io/api>. The perturbations include small molecules and other genetic perturbagen like shRNAs, cDNAs and biologics. They are treated on target genes for monitoring the possible downstream consequences of gene expression.

This project represents a small proof of concept database which integrates data from the CMap datasets with other publicly available biotechnology data in able to link the following data domains: genes, diseases, cell lines, tissues, publications, and perturbations. We hope to introduce to audience what CMap is and how should users manipulate the tools on it. It’s a meaningful platform to be used by biologists for not only discovering mechanism of action of new drugs, but also informing clinic trials. The analysis results via CLUE environment will definitely promote the development of therapeutic hypotheses.

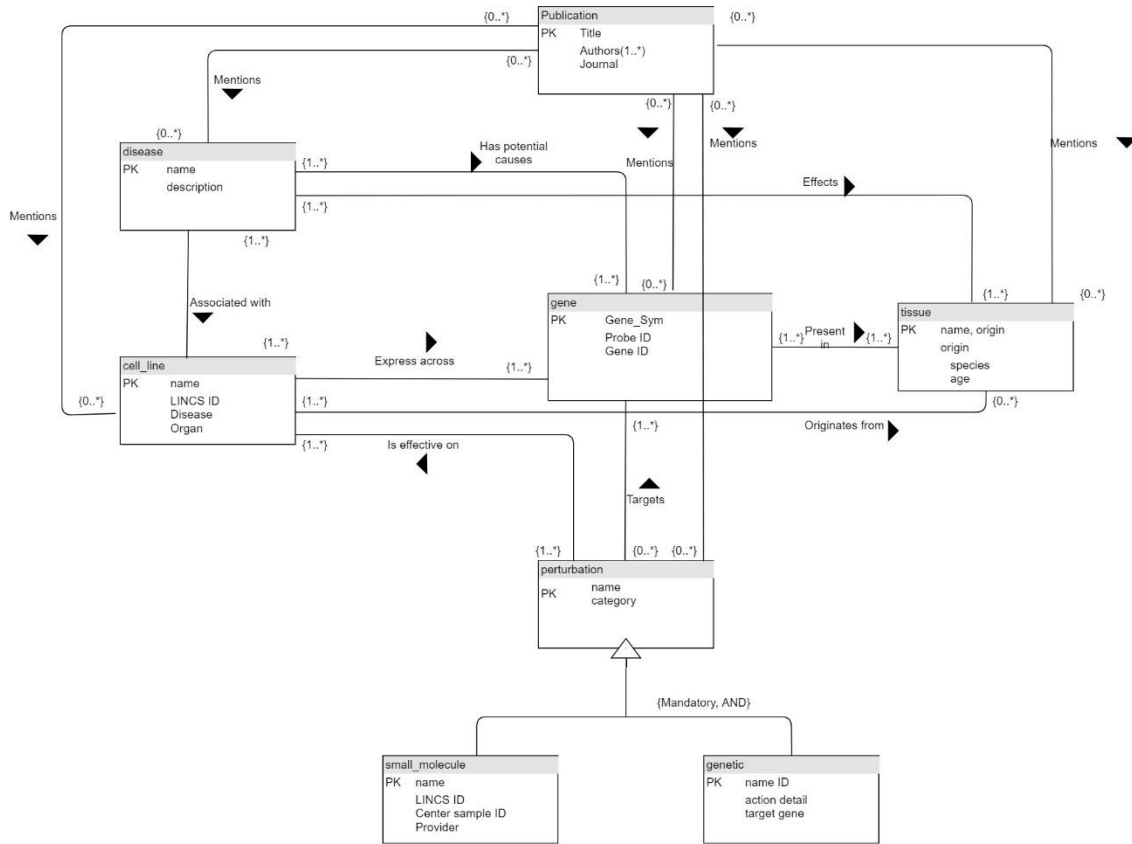
## 1.2 FUNCTIONALITY TO THE USER

database is designed to provide a tool for a researcher to add biological data to a database as a look-up table and find links among that data which may be useful for drug development or other correlational research. There should be two types of end user: 1) an administrator who can create/read/update/delete tuple data, and 2) a user who can only read. The two roles are corresponding to the backend and frontend users. Presumably, there will be an external system for submitting documents to review and insert. The user should be able to search by specific metrics, for example, gene, cell line, disease and literature.

## 1.3 STORAGE AND LOGISTICS

For the purpose of this project, we will be using SQL storage designed using MySQL Workbench. Metrics will be calculated and GUI designed in Python that can be run from the command line. There should not be any machine restrictions for the project other than the ability to run a Python program from the command line.

## 2 CONCEPTUAL DATABASE DESIGN



### 3 ACTIVITY DIAGRAM

